VEHICLE ACCIDENT SEVERITY

A description of the problem and background:

Every year the lives of approximately 1.35 million people are cut short as a result of a road traffic crash. Between 20 and 50 million more people suffer non-fatal injuries, with many incurring a disability as a result of their injury.

Road traffic injuries cause considerable economic losses to individuals, their families, and to nations as a whole. These losses arise from the cost of treatment as well as lost productivity for those killed or disabled by their injuries, and for family members who need to take time off work or school to care for the injured. Road traffic crashes cost most countries 3% of their gross domestic product. (Source: https://www.who.int/news-room/fact-sheets/detail/road-traffic-injuries)

In high income countries, the combination of safety standards and education have made tremendous strides in automobile safety. The challenge now is to ensure the low and middle income countries which now account for 50% of new car sales and production and over 90% of road fatalities can be brought up to global minimum standards.

As much knowledge as we have relating to vehicle safety, most still struggle to ensure they are doing everything necessary to guarantee their drivers are safe on the roadways and limit their exposure to crashes.

A description of the data:

The data set which will be used is the “Collisions—All Years” provided through the Applied Data Science Capstone by Coursera. They are data from SDOT Traffic Management Division.

The data consists of the information about the accidents. This includes all types of collisions from 2004 to present. The data has 194,673 rows and 38 columns. The columns are referred as the features of the dataset and every row referred in a single collision.

Our model built by features described below:

• SEVERITYCODE: specifies a particular code which corresponds to the severity of the collision:

2 –Injury
1 –Prop Damage

•VEHCOUNT: specifies the number of vehicles involved in a collision.

•JUNCTIONTYPE: specifies the category of junction at which the collision took place.

•WEATHER: specifies the weather condition at the time of collision.

•ROADCOND: specifies the condition of the road during the collision.

•LIGHTCOND: specifies the conditions of light during the collision.

I will explore the above features to discover if any of them have advanced possibilities to lead in injury when a collision occurs. Also, I'll try to build a model than can predict the severity of an accident to the people involved.

METHODOLOGY

The business problem:

Road traffic injuries can be prevented. Governments need to take action to address road safety in a holistic manner. This requires involvement from multiple sectors such as transport, police, health, education, and actions that address the safety of roads, vehicles, and road users.

Effective interventions can be done if we make good use of billions of collected data and the power of data science, building models which predict the severity of vehicles' accident.

Data understanding and preparation:

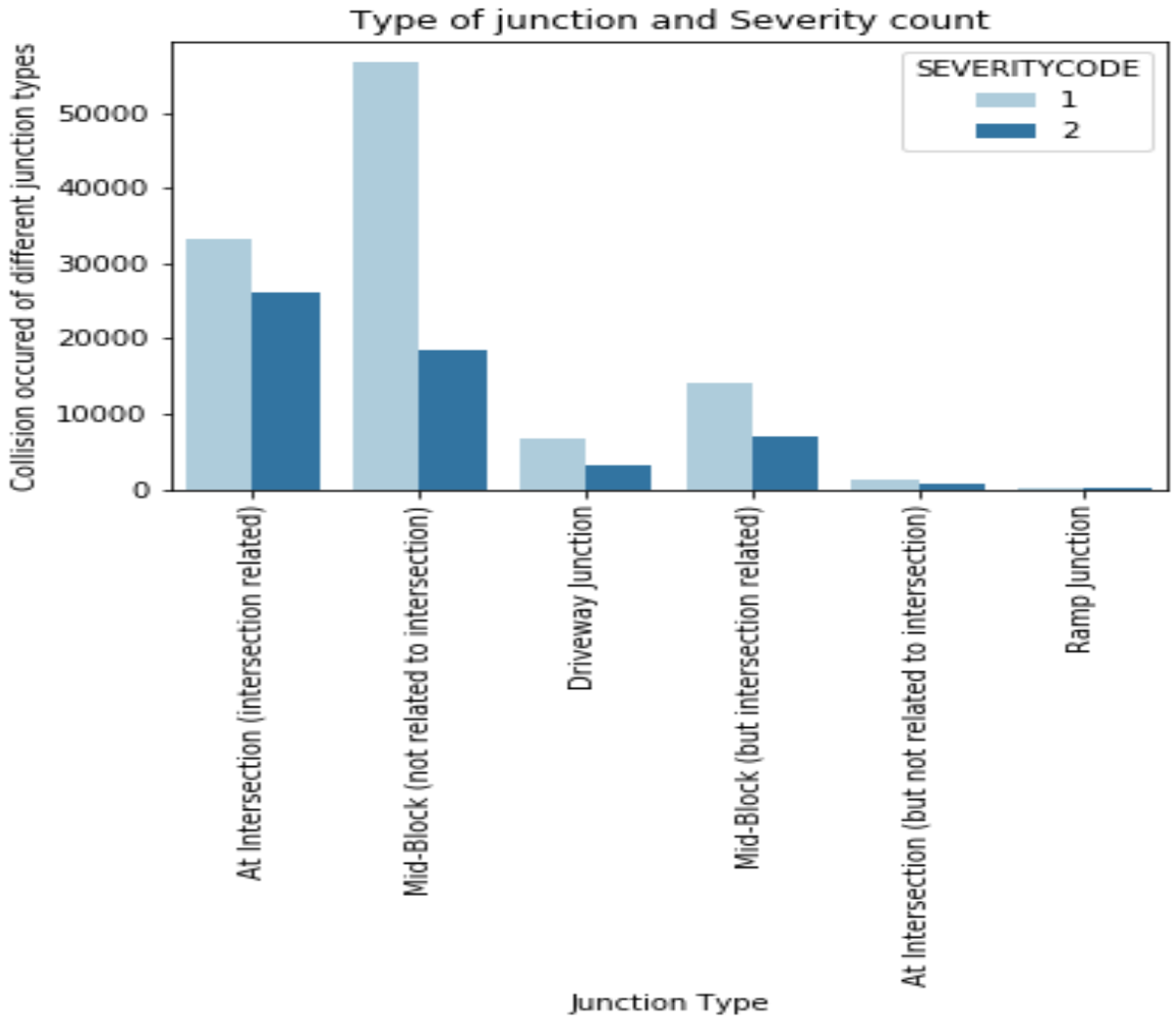
In this phase, we used the "Collisions—All Years" data set and we gave our attention to above mentioned features in order to explore their correlation with the severity of the collision.

We discover that there are some data that was described as “unknown” or “other” using simple functions and pandas for cleaning dataset.

Data analysis:

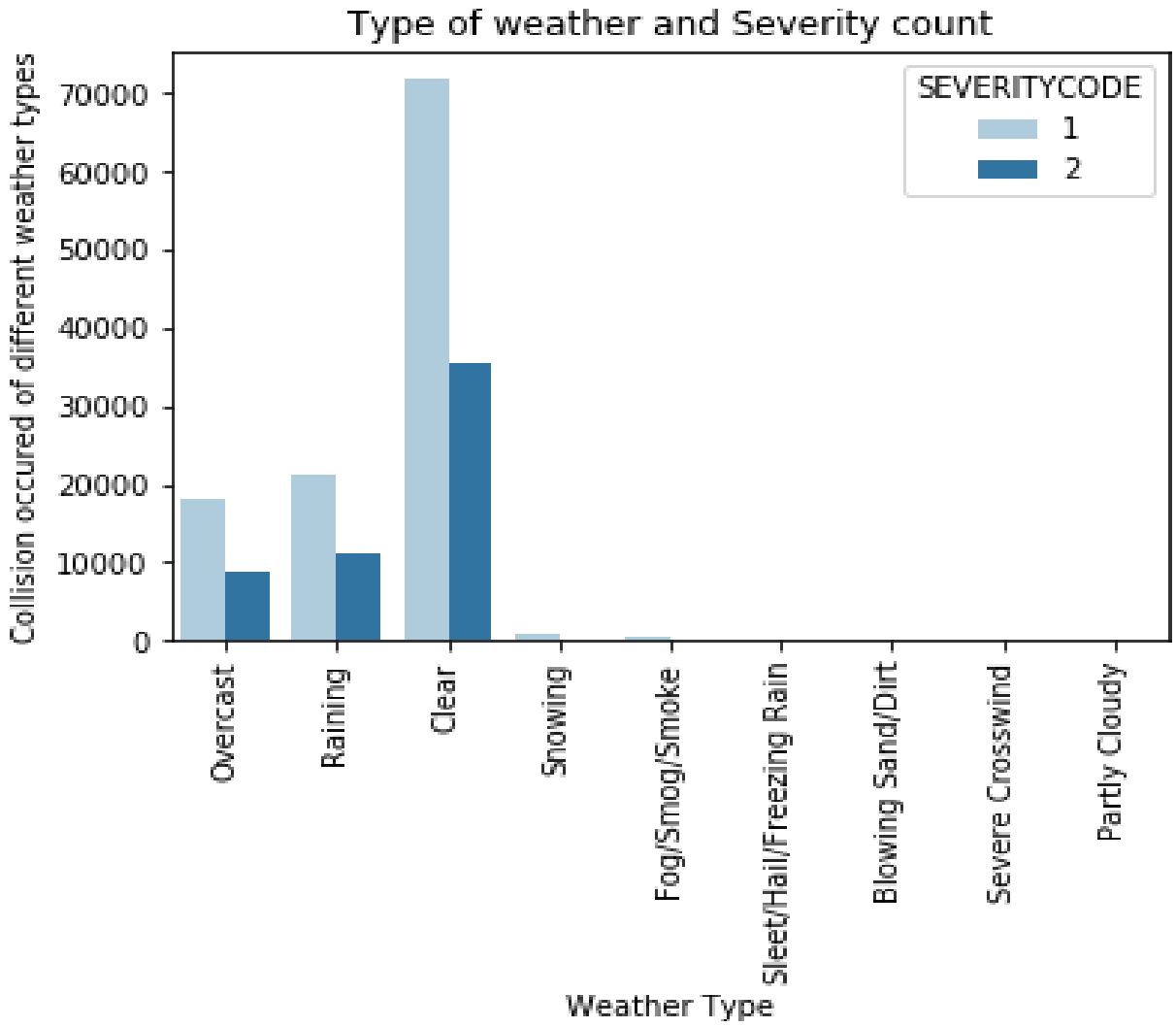
Junction type and severity

We can see that most Prop Damage accidents take place in Mid-Block (not related to intersection) while most accidents with injuries take place in At Intersection (intersection related).



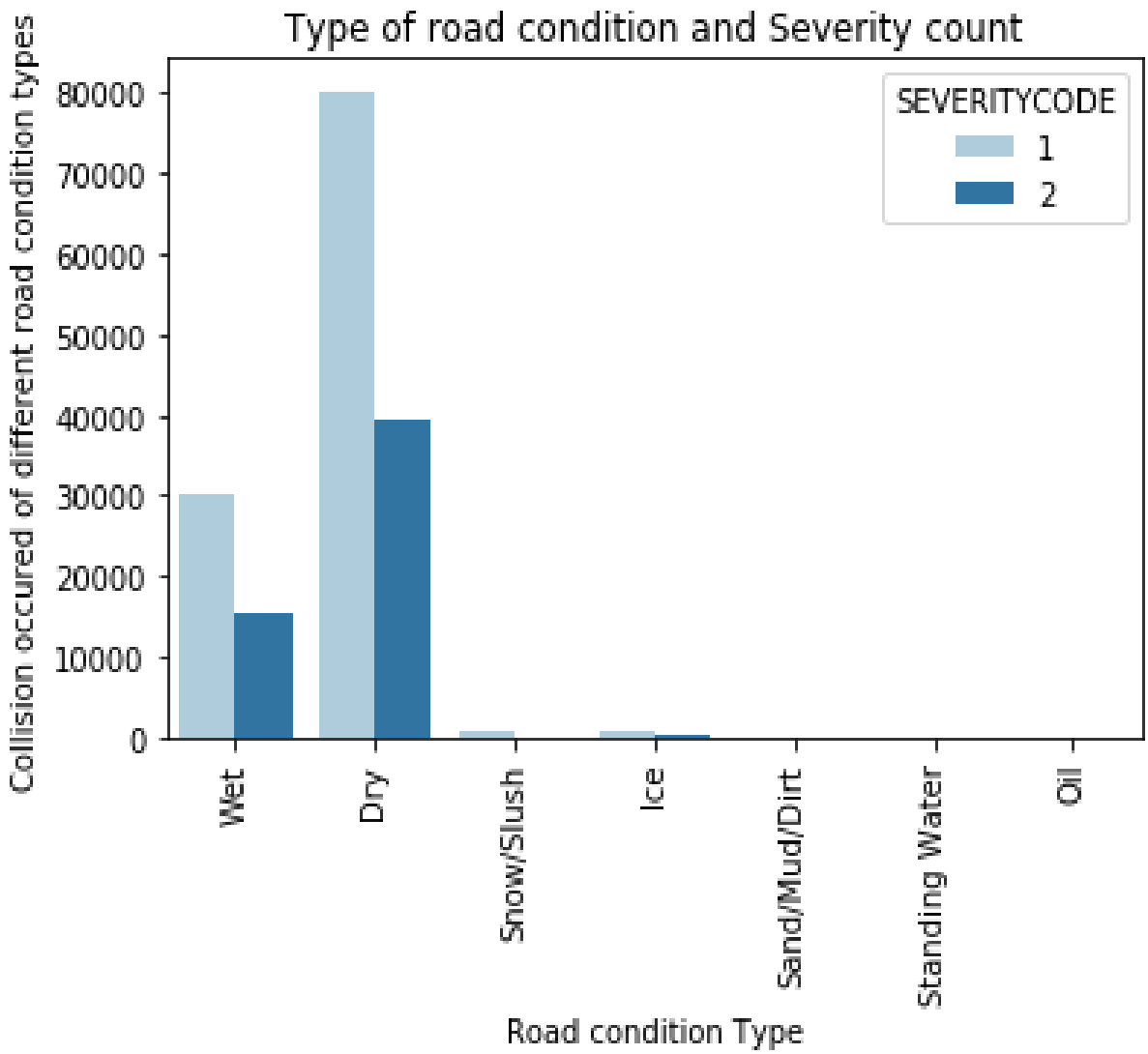
Weather and Severity

Surprising, majority of the accidents with injuries take place in good weather conditions.



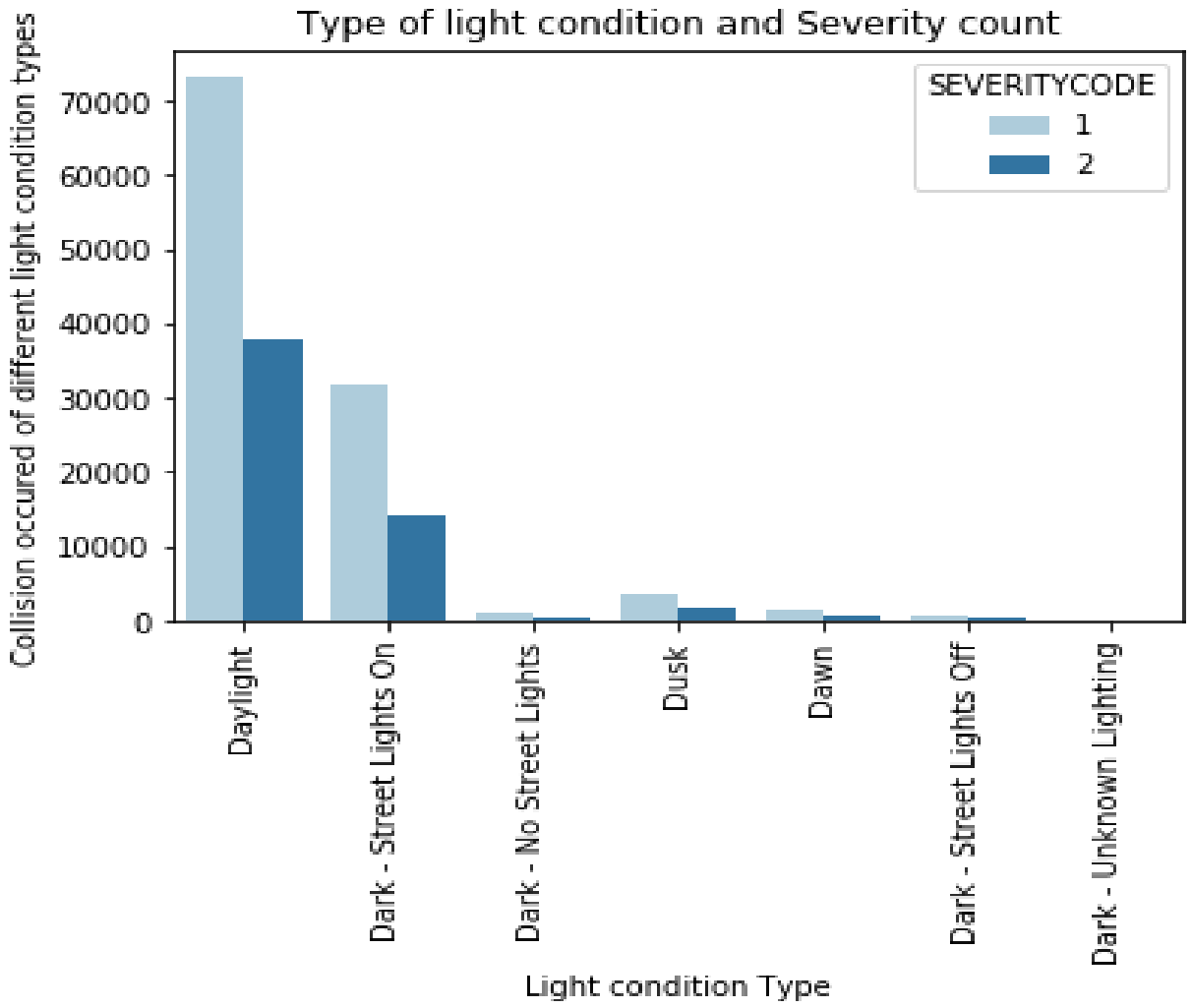
Road condition and severity

Most accidents take place in dry road conditions.



Light condition and severity

Only few accidents take place in poor visibility conditions. Most accidents take place in daylight.

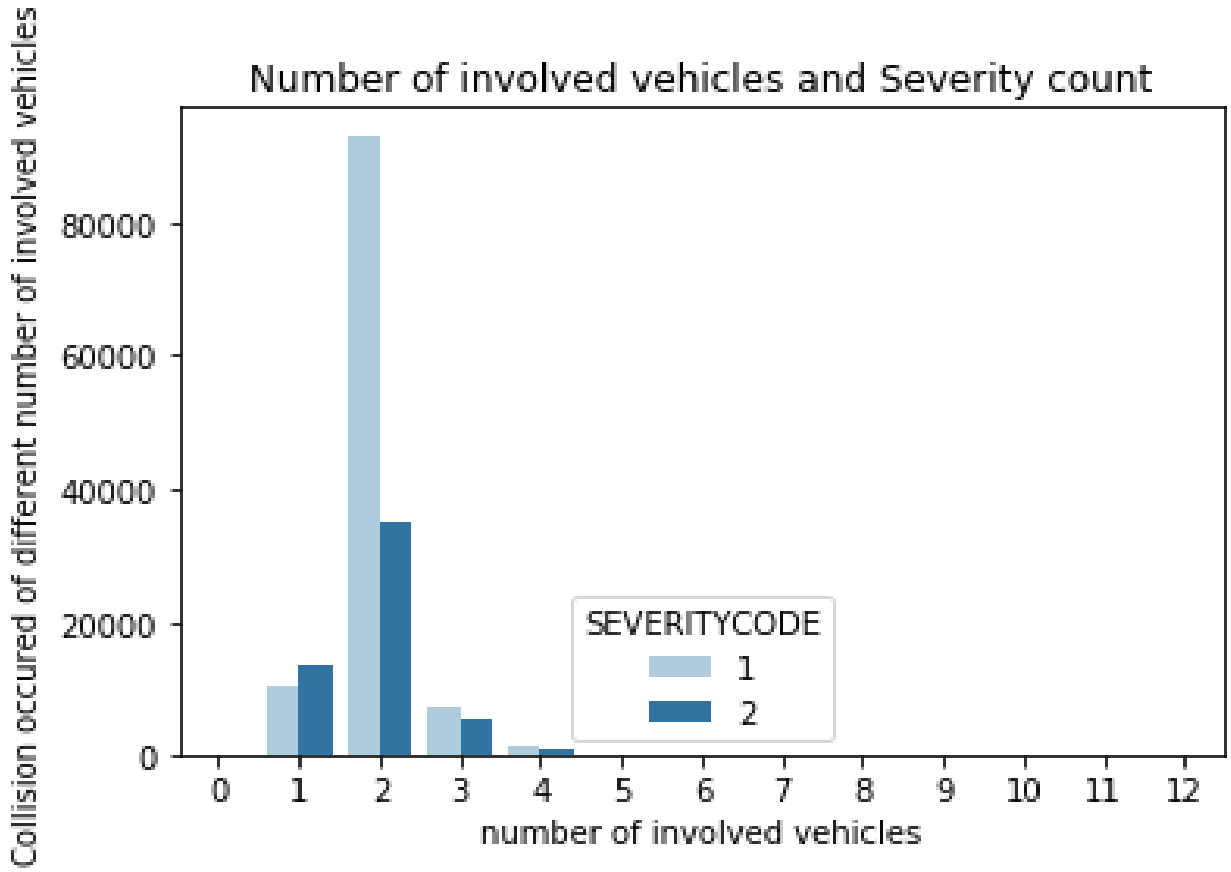


Light condition and severity

Only few accidents take place in poor visibility conditions. Most accidents take place in daylight.

Number of involved vehicles and Severity

In most accidents involved 2 vehicles. Anyway, multiple vehicle collisions are less often.



Modeling:

First we needed to convert the categorical columns of the selected features into numeric values so we can process them in our models.

In this phase, various algorithms and methods was selected and applied to build our model. We chose a rate of 80-20 on train and test our models.

The selected methods was :

- Logistic Regression
- Support Vector Machine
- Decision Tree
- Random Forest Classifier.

Evaluation:

The model needs to be evaluated thoroughly to ensure that the business objectives are achieved. Certain metrics used for the model evaluation.

We chose:

- Accuracy
- Recall
- F1-score
- Precision

Results:

- There are no significant correlation between the attributes and the severity of collision.
- The results of the different Machine Learning methods are the following:

MODEL	ACCURACY SCORE	F1 SCORE	PRECISION	RECALL
Logistic Regression	0.698522	0.815985	0.694297	0.989394
Decision Tree	0.708404	0.812444	0.718400	0.934818
Random Forest	0.707956	0.812195	0.718064	0.93473
Support Vector Machine	0.675593	0.806393	0.675593	1

- The best method to predict the severity of an accident is probably the Decision Tree method.

Discussion and Conclusion:

In our minds, there is tight correlation between weather conditions, road conditions, light conditions, junction types and severity of vehicles collisions. In reality there are so many attributes that affect the outcome of a crash so that the contribution of a specific one is too hard to be evaluated.

It remains the human influence the most significant factor for the result of an accident. The level of watchfulness, the driving experience and even the driving habits are difficult to be measured and to be part of a model.